BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 5: e75344
doi: 10.3897/biss.5.75344

OPEN ACCESS

Conference Abstract

A Set of Simple Tools For Assembling, Annotating, Versioning and Publishing Taxonomies

Laura Rocha Prado ‡

‡ Biodiversity Knowledge Integration Center, Arizona State University, Tempe, United States of America

Corresponding author: Laura Rocha Prado (laurarochaprado@gmail.com)

Received: 16 Sep 2021 | Published: 16 Sep 2021

Citation: Rocha Prado L (2021) A Set of Simple Tools For Assembling, Annotating, Versioning and Publishing Taxonomies. Biodiversity Information Science and Standards 5: e75344. https://doi.org/10.3897/biss.5.75344

Abstract

Biodiversity data publishers rely on virtually assembled taxonomic hierarchies to structure their data, with operational units involving scientific names, nomenclatural acts and taxonomic trees. The main goal for the majority of biodiversity aggregators, databases, and software developed specifically for managing scientific names, biological samples and other occurrences has been to establish a single, unified biological classification, to serve as their structural "taxonomic backbone." Resources to produce and publish biological classifications digitally are thus, typically restricted to those generating unified taxonomic backbones, leaving individual researchers and decentralized communities with few options to assemble, visualize, version and disseminate multiple taxonomies online.

To aid the creation of a culture of assembling, annotating, versioning, and publishing taxonomies online, and to help users interested in taxonomic classifications that lack digital communities, the development of a set of modular and independent tools is proposed, based on the following complementary features:

- A web application to serve as the taxonomy curator (referred to as the Curator)
- A web application to serve as the optional taxonomic database and information provider (referred to as the Aggregator)

These tools are being designed and built following modern software development standards, in a modular architecture consisting of front-end clients, databases, and back-

© Rocha Prado L. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

end applications, with the provision for a public Application Programming Interface (API) that will make data available for any interested parties and can be potentially integrated into large-scale projects like the Global Biodiversity Information Facility (GBIF), Integrated Digitized Biocollections (iDigBio), Symbiota (Gries et al. 2014), and Plazi (Agosti and Egloff 2009).

Curator tool

The Curator tool will be a publicly accessible front-end web application, with which users can assemble, curate, and export taxonomies. The primary focus is to support the user-preferred taxonomy generation, with manual inputs and optional annotations of the resulting product. Users can pick between three modes of taxonomy assembly:

1. manual mode with assisted taxon search,
2. automated generation from an online source, and
3. automated generation from a file upload. Taxonomies can be edited and annotated as necessary.

Once a user is satisfied with their taxonomy, they can save it in one or all of the available formats for exporting and external usage (common formats include, among others, JSON (JavaScript Object Notation), CSV (comma-separated values), and XML). Logged in users can also opt to save the taxonomy in the Aggregator database, which will make the taxonomy publicly available. Ideally, all fields in the Curator forms should correspond to terms included in the Darwin Core standard (Wieczorek et al. 2012) or Plazi's TaxonX schema (Agosti and Egloff 2009) (for hierarchies available in published treatments).

Aggregator tool

The Aggregator tool will communicate with the database and will provide users with a number of functionalities, such as:

- Store and publish versioned taxonomies generated with the Curator
- API endpoints for automation (JSON/XML formats/CSV download)
- Optional unique identifier/DOI generation for published taxonomies
- Search engine with user-friendly interface as well as API endpoint for querying the database

The possibility of making taxonomies available as an API endpoint, as well as exporting taxonomies in different formats, will ensure that this tool behaves as a taxonomic source that can be used by virtually any interested party or application. The tools are being modelled as a decentralized community resource that can be used for any or all taxonomic groups and, as such, its scale and impact will be driven by bottom-up community use. The goal is not to provide extensive coverage of all biological organisms, but rather to provide an open digital toolkit and space for biodiversity researchers and projects that lack access to open, structured, online taxonomic publication venues and dedicated tools.

Practical examples of usage for these tools include:

- A user generates multiple taxonomic concepts for organisms they are studying, which can then be queried and analyzed by scripts that make taxonomic alignments to compare different scientific hypotheses throughout time;
- An institution wants to publish a regional Symbiota portal to manage specimens in a particular collection, so they establish an annotated working taxonomic backbone with the Curator that Symbiota will then be able to ingest before samples can be imported into the portal;
- A researcher wants to export a biodiversity portal taxonomy at a given moment and wants to annotate and publish this version in an upcoming paper to establish scientific baselines for proper taxonomic communication.

Keywords

biological classifications, cybertaxonomy

Presenting author

Laura Rocha Prado

Presented at

TDWG 2021

References

- Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2 https://doi.org/10.1186/1756-0500-2-53
- Gries C, Gilbert E, Franz N (2014) Symbiota – A virtual platform for creating voucher-based biodiversity information communities. Biodiversity Data Journal 2 https://doi.org/10.3897/BDJ.2.e1114
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715